J. Rauhala, H.-M. Lehtonen, and V. Välimäki, Fast automatic inharmonicity estimation algorithm, Journal of the Acoustical Society of America, vol. 121, no. 5, pp. EL184-EL189, 2007.

Fast automatic inharmonicity estimation algorithm

Jukka Rauhala, Heidi-Maria Lehtonen, and Vesa Välimäki
Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, P.O. Box 3000, FI-02015 TKK, Espoo, Finland.
jukka.rauhala@acoustics.hut.fi, heidi-maria.lehtonen@tkk.fi, vesa.valimaki@tkk.fi

Abstract: A new algorithm is presented for estimating the inharmonicity coefficient of slightly inharmonic stringed instrument sounds. In the proposed partial frequencies deviation method, the inharmonicity is estimated in an intuitive way by minimizing the deviation of the expected partial frequencies compared to the frequencies of the high amplitude peaks in the spectrum. This is done in an iterative process, where the algorithm converges towards the target estimation value. The algorithm is tested using both synthetic and recorded piano tones. The results show that the new algorithm produces accurate results with a small computation cost compared to other methods.

PACS numbers: 43.75.Yy, 43.75.Mn [DKW]
Date Received: November 20, 2006 **Date Accepted:** February 23, 2007

1. Introduction

Inharmonicity is a phenomenon occurring in string instruments due to the stiffness of the string and nonrigid terminations, and it may have an audible effect on the sound produced by the instrument [Järveläinen *et al.*, 2001]. For example, it is generally agreed that the characteristic timbre of the piano is caused in part by the inharmonicity [Fletcher *et al.*, 1962]. The value of the inharmonicity coefficient must be estimated from real data in order to study the auditory perception of the phenomenon or to synthesize these instrument sounds accurately [Bank *et al.*, 2003; Bensa *et al.*, 2005].

The estimation of the inharmonicity value out of recorded data has two main challenges. First, the estimation of the fundamental frequency is required, which can be a difficult task. Second, it is challenging to extract the partial frequencies from real data, because there can be high spectral peaks in the data (due to, e.g., longitudinal modes, or nonlinearities) which do not belong to a single series of partial frequencies. For example, most of the keys of the piano have three strings with slightly different fundamental frequency and inharmonicity values.

A few previous studies have tackled inharmonicity estimation. Galembo and Askenfelt have developed a method based on an inharmonic comb filter realized in the frequency domain to estimate the inharmonicity value [Galembo and Askenfelt, 1999]. Also, cepstral analysis and the harmonic product spectrum, both popular pitch extraction techniques, have been used to estimate inharmonicity [Askenfelt and Galembo, 2000]. Another method for inharmonicity estimation is based on estimating the fundamental frequency in subbands [Klapuri, 2003].

2. Estimation algorithm

Two requirements are defined for the estimation algorithm. First, it should be fully automatic; it should not require parameters other than the fundamental frequency. Second, it should be able to estimate the inharmonicity value from real data with a significant amount of noise, a common problem with recordings of acoustic musical instruments.

Inharmonicity means that the partial frequencies appear to be higher than those of the ideal string. The frequencies of inharmonic partials can be calculated as [Fletcher *et al.*, 1962]

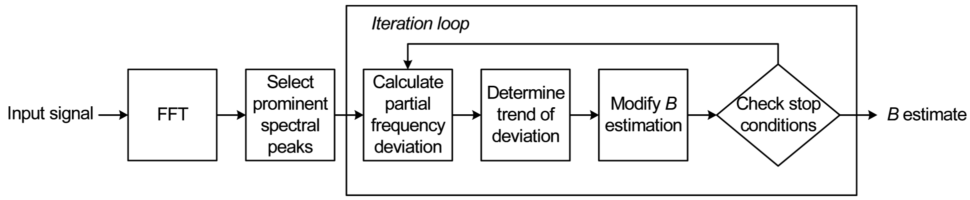

Fig. 1. Structure of the proposed PFD algorithm.

$$f_k = kf_0\sqrt{1 + Bk^2}, \tag{1}$$

where k is the partial number, f_0 is the nominal fundamental frequency of an ideal (nondispersive) string, and B is the inharmonicity coefficient.

Instead of using f_0, which is a theoretical value not existing in the spectrum, the first partial f_1 present in the spectrum is used to calculate the inharmonicity values. Note that when f_0 is substituted with $f_1/(1+\mathrm{B})^{-0.5}$ in Eq. (1), f_0 need not be calculated in order to compute B. This reduces the number of variables to one without affecting the results. A simple technique based on the fast Fourier transform (FFT) is used for estimating f_1, because common pitch detection algorithms, such as YIN [de Cheveigné and Kawahara, 2002], do not work well for inharmonic signals. In this technique, the frequency of the highest amplitude in the spectrum within a given range (see Sec. III) is picked, and the final estimation is done by using three-point parabolic interpolation.

An intuitive way to search for the B value manually would be to plot the spectrum of the signal and the partial frequencies corresponding to the estimated B value. This would be done iteratively, and the B estimate would be modified at each iteration according to the displacement between the partial frequencies and high amplitude peaks in the spectrum to minimize their difference. The algorithm proposed in this letter mimics this method in an automatic way.

Figure 1 shows the block diagram of the proposed algorithm. The first step is to determine the spectrum of the data by calculating the FFT of a long sample from the starting point of the signal. The data are windowed with the Blackman window. The number of FFT points used, 2^{16} points after zero padding in this work, should be large for good resolution. In the second step, the number of the spectral peaks is reduced using a simple technique, where the spectrum is divided into subbands and ten spectral peaks with the highest amplitudes selected from each subband. The width of the subbands is defined to be $5f_1$. Hence, the number of selected peaks is reduced approximately to $2K_{\max}$, where $K_{\max}$ is the maximum number of partials used in the estimation.

The main part of this algorithm is the iteration loop, which consists of three parts: partial frequencies deviation calculation, determination of the deviation trend, and B estimate (denoted as $\hat{B}$) modification. The initial value of $\hat{B}$ should be within the range of typical B values (10^{-4} to 10^{-3} in piano bass tones), but it does not need to be very accurate as it does not affect much the final result. In this work, the initial value used is $B=10^{-4}$. In the first part, $\hat{B}$ is used for calculating estimated partial frequencies $\hat{f}_k$ via Eq. (1). In order to avoid bias due to the limited frequency resolution, $\hat{f}_k$ are quantized using the frequency resolution due to FFT. Then, the frequencies f_k corresponding to the highest amplitudes within the closed interval $[\hat{f}_k-\Delta f,\hat{f}_k+\Delta f]$ are selected, and the partial frequency deviation $D_k=\hat{f}_k-f_k$ is calculated. In this work, the parameter $\Delta f=0.4f_1$.

The next task in the iteration loop is to determine the trend of the deviation which indicates whether $\hat{B}$ should be decreased or increased. The problem is that the deviation curve is not always smooth and there can be sharp peaks and notches in the curve. This is taken into account by defining the trend to be positive, if there are more positive derivates than negative

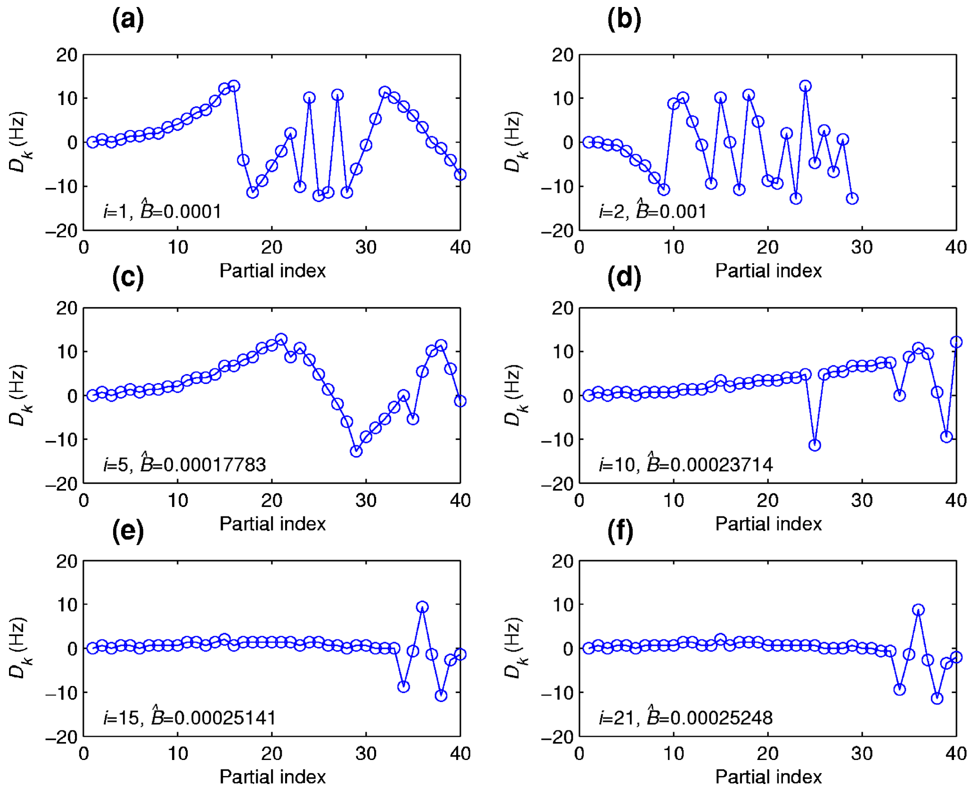

Fig. 2. (Color online) The partial frequency deviation D_k curve at iteration loop rounds (a) 1, (b) 2, (c) 5, (d) 10, (e) 15, and (f) 21, when a recorded Bb_0 tone is analyzed (f_0=32.4 Hz). The corresponding $\hat{B}$ values at each round are shown in the figures, the final B estimate is 0.0002546.

derivates, and vice versa. Hence, the trend can be calculated, for example, as $\mathrm{sign}(\Sigma_{k=2}^{K_{\max}}\mathrm{sign}(D_{k+1}-D_k))$. The latter part of the iteration is the modification of $\hat{B}$, based on the determined trend of the deviation. The $\hat{B}$ value is changed by multiplying it by 10^{δ}, where δ is an adaptive step-size parameter (the initial value is set to 1 in this work), which has a positive sign if the trend is positive, or a negative sign if it is negative. Whenever a reverse occurs (that is, two adjacent trends have different signs), parameter δ is modified by dividing it by 2. In other words, the goal is to increase (or decrease, depending on the sign of the determined trend) $\hat{B}$ until it is larger than the target B estimate, which can be seen in the D_k curve as a change of trend. Then, the iteration loop is performed towards the opposite direction with a decreased step size, assuming that the iteration loop converges around the target B estimate value. Figure 2 and Mm. 1 illustrate how the deviation curve behaves in different stages of the iteration loop. The trade-off between the speed of the algorithm and the accuracy of the algorithm can be controlled by the loop stop conditions. This work uses the following stop conditions: the iteration loop is run until either it has been run for 40 times or $|\delta| < 10^{-4}$. Finally, the proposed method provides a way to evaluate the produced B estimate reliably, as the smoothness of the last D_k curve indicates the level of accuracy.

Mm. 1. Algorithm convergence with a synthetic signal (1.3 MB). This is a file of type “mpeg.”

In addition to B estimation, the D_k curve provides information on the accuracy of the f_1 estimate: convexity in the curve indicates that the estimate should be increased and concavity indicates the need to decrease the estimate. This can be used for refining the f_1 estimate, which improves the accuracy of B estimation as well. The B estimation by using the f_1 estimate refining is done as follows. First, the B estimation is done as proposed above to obtain the initial B estimate value. Then, the same algorithm, as shown in Fig. 1, is used for f_1 estimate refining, only this time instead of analyzing the trend of the deviation and modifying B, the convexity and concavity of the deviation is estimated, and the f_1 estimate is modified. The convexity and concavity can be estimated by examining the average of the first half of D_k values: the positive curve is convex and vice versa. Based on this, the f_1 estimate is modified by multiplying it with $(1+\mu)$, where μ is the step size, which has a positive sign if the curve is convex and a negative sign if the curve is concave. The initial value of μ is set to 0.005. Similarly to δ, μ is divided by 2 at each change in trend. The refining iteration is stopped when it has been run for 100 times or $|\mu| < 10^{-5}$. After the f_1 estimate is refined, the original B estimation algorithm is run again to modify the B estimate.

3. Results and comparison

The proposed algorithm was compared against two versions of the inharmonic comb filter (ICF) method [Galembo and Askenfelt, 1999]. In the ICF method, which was included in the comparison because it is a good previously introduced method, the signal is filtered with a multiband filter. The center frequencies of the pass bands are determined from Eq. (1). The ICF method does not need an external f_0 estimation method, as it combines the inharmonicity estimation and the f_0 estimation.

In the first phase of the ICF method, a rough estimation of f_0 is made with a pre-defined value for the inharmonicity. Then, the rough f_0 estimate is used to calculate a rough estimation for the inharmonicity value. Both f_0 and B values are varied and a power spectrum sum is calculated at each point. The estimated inharmonicity value is determined to be the value corresponding to the maximum power spectrum sum. In the final phase, the ranges of the B and f_0 values are narrowed down around the rough estimates obtained in the previous phases. At the end, the result from the third phase is refined by parabolic interpolation. The details of the algorithm are documented in [Galembo and Askenfelt, 1999].

In order to improve the performance of the ICF method in the tests, the parameters suggested in [Galembo and Askenfelt, 1999] were slightly modified (the modified version is denoted in this work as ICF+). First, the bandwidth of the pass band was chosen to be 5% of f_0. Second, a linear scale was used in the second phase of the algorithm to produce better results than with the logarithmic scale proposed in [Galembo and Askenfelt, 1999]. Moreover, the number of the B values on the grid in the second phase was changed to 400, and in the third phase it was chosen to be 100 according to test in this work.

The number of partials included in the estimation was chosen to be 50 for the proposed algorithm, whereas for the ICF and ICF+ it was 30, as suggested in [Galembo and Askenfelt, 1999]. The sampling rate used was 44,100 Hz in all cases. Synthetic signals generated with additive synthesis were used in the first case. The correct fundamental frequency used in synthesis was given to all methods in order to avoid the bias caused by the fundamental frequency extraction. The second case tested the inharmonicity estimation with real piano data. In this case, each method used its own fundamental frequency estimation to determine the fundamental frequency from the data (PFD was run with the f_1 estimate refining, as described in Sec. II). Since the piano key number is known in all cases, the methods assumed that the correct f_0 lies within the range a major third up and down from f_0 (based on an equal-tempered scale with A_4=440 Hz) in order to prevent octave errors. The test cases were limited to key range 1–35, since the original ICF method is designed for $f_0 < 200$ Hz. Moreover, the inharmonicity effect is most important in the bass range of the piano [Galembo and Askenfelt, 1999].

The first test case simulated the inharmonicity estimation of the piano sounds. A synthetic signal was created for each key of the piano. A typical inharmonicity value of the piano was selected for each signal, see, e.g., [Askenfelt and Galembo, 2000]. The partial amplitudes were measured from real data, with some manual fine tuning, and a variation of 0–10% was added to the synthetic partial amplitudes. Each key included the correct number of synthesized

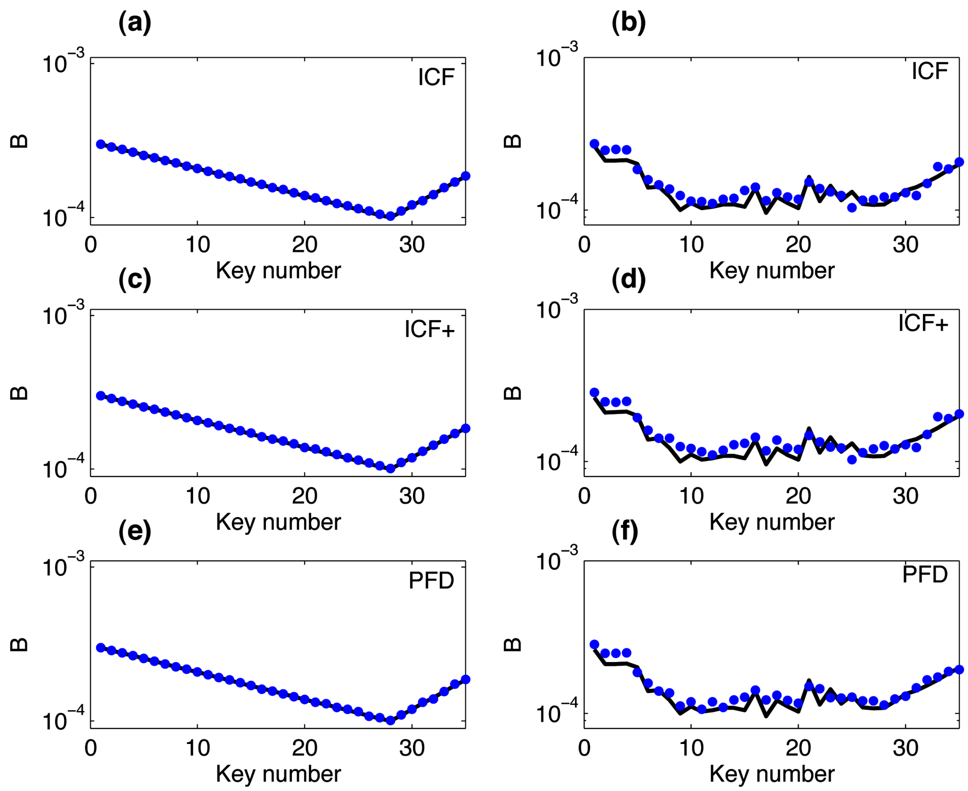

Fig. 3. (Color online) Inharmonicity estimation test for synthetic piano data with SNR of 40 dB using (a) ICF, (c) ICF+, and (e) PFD. The solid line represents the correct inharmonicity coefficient values for keys 1-35. The same test was done with recorded piano tones[1] by using (b) ICF, (d) ICF+, and (f) PFD. The solid line indicates manually estimated values.

strings, i.e., keys 1–12 had only one string, whereas keys 13–29 had two, and keys 30–35 had three strings. The signal-to-noise ratio (SNR) was determined by comparing the power of the signal during its first 10 ms to the power of the noise in a 10 ms window. Gaussian white noise was added into the signals so that a SNR of 40 dB was obtained during the first 10 ms of the signal. The test signals are available at http://www.acoustics.hut.fi/publications/papers/pfd. Finally, the three methods were tested by using real piano samples.[1] The SNR in these piano tones was approximately 35 dB.

Figure 3 shows the test results for the two test cases. In addition, the total running times and average rms error values corresponding to the first test case and average rms deviation values for the test case using recorded tones are shown in Table 1. The results from the first test case, which used synthetic piano tones, show that all methods produce good results. The differences are very small and the methods can be considered equal in terms of the rms error. In the second test case, PFD had the best performance according to the rms values. Moreover, ICF and ICF+ produced slightly less accurate results. The differences in the computational cost between the presented method and the inharmonic comb filter method variations were remarkable: the running time of the test case using the real piano tones for ICF+ was over 15 min, whereas PFD produced better results in less than 1 min. Hence, the proposed method was proven to produce good results in a fraction of the time compared to other methods.

4. Conclusion

An algorithm for estimating the inharmonicity coefficient for the sounds of stringed musical instruments was presented. The algorithm is simple and it is able to produce fairly accurate

Table 1. Total running times of test cases, and average RMS errors (synthetic tones) and average RMS deviations (real tones) of the estimated inharmonicity values for the three methods.

	Synthetic tones		Real tones	
Method	RMS error	Running time	RMS deviation[a]	Running time
ICF	1.19×10^{-6}	11.4 s	1.73×10^{-5}	313.9 s
ICF+	1.16×10^{-6}	24.1 s	1.88×10^{-5}	946.5 s
PFD	1.19×10^{-6}	7.0 s	1.66×10^{-5}	58.4 s

[a]Difference from manually estimated values. The correct inharmonicity value is unknown.

estimations at a small computational cost. In addition, it is able to refine fundamental frequency estimation. The proposed algorithm can be used as an analysis tool to obtain a good estimate of the inharmonicity of various string instruments. Future work includes examination of improving the accuracy of the B estimation, and study on using the algorithm for f_0 estimation with stringed instrument tones.

Acknowledgment

This work was financially supported by the Academy of Finland (Project No. 104934). J. R. was supported by the Nokia Foundation, and H.-M. L. was supported by Tekniikan edistämissäätiö and the Emil Aaltonen Foundation.

References and links

[1]Steinway grand piano samples from University of Iowa (http://theremin.music.uiowa.edu Viewed 4/2/07).

Askenfelt, A., and Galembo, A. S. (**2000**). "Study of the spectral inharmonicity of musical sound," Acoust. Phys. **46**(2), 121–132.

Bank, B., Avanzini, F., Borin, G., De Poli, G., Fontana, F., and Rocchesso, D. (**2003**). "Physically informed signal processing methods for piano sound synthesis: A research overview," EURASIP J. Appl. Signal Processing **2003**(10), 941–952.

Bensa, J., Bilbao, S., Kronland-Martinet, R., Smith III, J. O., and Voinier, T. (**2005**). "Computational modeling of stiff piano strings using digital waveguides and finite differences," Acta. Acust. Acust. **91**(2), 289–298.

de Cheveigné, A., and Kawahara, H. (**2002**). "YIN, a fundamental frequency estimator for speech and music," J. Acoust. Soc. Am. **111**(4), 1917–1930.

Fletcher, H., Blackham, E. D., and Stratton, R. (**1962**). "Quality of piano tones," J. Acoust. Soc. Am. **34**(6), 749–761.

Galembo, A. S., and Askenfelt, A. (**1999**). "Signal representation and estimation of spectral parameters by inharmonic comb filters with application to the piano," IEEE Trans. Speech Audio Process. **7**(2), 197–203.

Järveläinen, H., Välimäki, V., and Karjalainen, M. (**2001**). "Audibility of the timbral effects of inharmonicity in stringed instrument tones," Acoust. Res. Letters Online, **2**(3), 79–84.

Klapuri, A. (**2003**). "Multiple fundamental frequency estimation based on harmonicity and spectral smoothness," IEEE Trans. Speech Audio Process. **11**(6), 184–194.